DOCUMENT RESUME

ED 085 637 CG 008 578

AUTHOR Flammer, August
TITLE A Conceptual Model Relating Transfer of Learning and Correlation: Theoretical Paper No. 43.
INSTITUTION Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Jul 73
CONTRACT NE-C-00-3-0065
NOTE 28p.; Report from the Project on Cognitive Operations and Abilities in Concept Learning

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Academic Achievement; *Correlation; Educational Theories; *Learning Theories; *Models; *Performance Factors; Psychometrics; *Transfer of Training

ABSTRACT

Discussed is the theoretical and practical importance of the development of learning and transfer taxonomies with psychometric relevance and of the building of psychometric classificatory systems with implications for learning and instruction. Psychometric classifications of human performances most often are based on the covariation of individual differences. A model is presented which justifies the expectation that the transfer from learning one task to learning another is linearly dependent on the coefficient of intercorrelation between the two tasks when the coefficient is corrected for attenuation. Two studies so far have explicitly confirmed the main deductions from this model. Contrary to the predictions, however, the regression curves yielded negative intercepts. Two empirically testable explanations are offered, one of which would be in full accordance with the model, while the other would call for a further assumption. (Author)

FILMED FROM BEST AVAILABLE COPY

ED 085637

THEORETICAL PAPER NO. 43

A CONCEPTUAL MODEL RELATING TRANSFER OF LEARNING AND CORRELATION

REPORT FROM THE PROJECT ON COGNITIVE OPERATIONS AND ABILITIES IN CONCEPT LEARNING

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

WISCONSIN RESEARCH AND DEVELOPMENT
CENTER FOR COGNITIVE LEARNING

THE UNIVERSITY OF WISCONSIN
Madison, Wisconsin

National Institute of Education
Center Contract No. NE-C-00-3-0065

CE 008 578

RESEARCH COGNITIVE LEARNING DEVELOPMENT

ED 085637

Theoretical Paper No. 43

A CONCEPTUAL MODEL RELATING TRANSFER OF LEARNING AND CORRELATION

by

August Flammer

Report from the Project on Cognitive Operations
and Abilities in Concept Learning

Herbert J. Klausmeier
Principal Investigator

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

July 1973

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the National Institute of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by that agency should be inferred.

Center Contract No. NE-C-00-3-0065

Statement of Focus

Individually Guided Education (IGE) is a new comprehensive system of elementary education. The following components of the IGE system are in varying stages of development and implementation: a new organization for instruction and related administrative arrangements; a model of instructional programing for the individual student; and curriculum components in prereading, reading, mathematics, motivation, and environmental education. The development of other curriculum components, of a system for managing instruction by computer, and of instructional strategies is needed to complete the system. Continuing programmatic research is required to provide a sound knowledge base for the components under development and for improved second generation components. Finally, systematic implementation is essential so that the products will function properly in the IGE schools.

The Center plans and carries out the research, development, and implementation components of its IGE program in this sequence: (1) identify the needs and delimit the component problem area; (2) assess the possible constraints—financial resources and availability of staff; (3) formulate general plans and specific procedures for solving the problems; (4) secure and allocate human and material resources to carry out the plans; (5) provide for effective communication among personnel and efficient management of activities and resources; and (6) evaluate the effectiveness of each activity and its contribution to the total program and correct any difficulties through feedback mechanisms and appropriate management techniques.

A self-renewing system of elementary education is projected in each participating elementary school, i.e., one which is less dependent on external sources for direction and is more responsive to the needs of the children attending each particular school. In the IGE schools, Center-developed and other curriculum products compatible with the Center's instructional programing model will lead to higher student achievement and self-direction in learning and in conduct and also to higher morale and job satisfaction among educational personnel. Each developmental product makes its unique contribution to IGE as it is implemented in the schools. The various research components add to the knowledge of Center practitioners, developers, and theorists.

Acknowledgments

This theoretical paper is a further elaborated version of a paper presented at the 1973 American Psychological Association Convention in Montreal, Canada. The work has been done under a fellowship by the Swiss National Fund of Scientific Research (Grant SG 86) while the author was an Honorary Fellow at the Wisconsin Research and Development Center for Cognitive Learning at the University of Wisconsin in Madison. Both are gratefully acknowledged. Further thanks is expressed to the members of a former research team that produced the data on which part of the experimental test of the proposed model is based: Rudolf Flühler, Jo Kramis, Urs Murer, and Heinz Stöckli, all at the University of Fribourg in Switzerland.

The author also wishes to thank Diane H. Eich for competent correction and edition of this second-language report.

Table of Contents

		Page
	Acknowledgments	iv
	List of Tables	vii
	Abstract	ix
I.	Problem	1
	Early Studies	1
II.	A Conceptual Model	5
III.	Validation Experiment	9
	Subjects	9
	Learning task	9
	Transfer tasks	9
	Experimental training	10
	Transfer scaling	10
	Hypotheses	10
	Results	10
IV.	Discussion	13
	References	15

List of Tables

Table		Page
1	Linear Correlations r_{TC} Between Transfer Measures and Correlations	3
2	Illustrative Examples of Transfer of Learning	8
3	Correlations Between Predictor-Correlations and Transfer Measures	11
4	t-Values for Departure from Zero of the Intercepts of the Linear Regression Curves	12

Abstract

The theoretical and practical importance of the development of learning and transfer taxonomies with psychometric relevance and of the building of psychometric classificatory systems with implications for learning and instruction is discussed. Psychometric classifications of human performances most often are based on the covariation of individual differences. A model is presented which justifies the expectation that the transfer from learning one task to learning another is linearly dependent on the coefficient of intercorrelation between the two tasks when the coefficient is corrected for attenuation. Two studies so far have explicitly confirmed the main deductions from this model. Contrary to the predictions, however, the regression curves yielded negative intercepts. Two empirically testable explanations are offered, one of which would be in full accordance with the model, while the other would call for a further assumption.

I
Problem

Behavioral performance relies heavily on prior learning. Such learning often affects different kinds of performance at the same time, as a learning activity is not usually undertaken with the intention to produce just one "discrete" performance in one "discrete" situation. This calls for transfer. If a learning activity is really to be effective, its result has to be transferable, since every performance situation differs in some manner from every other.

Although the reality of the transfer phenomenon is undeniable, there seems not to be transfer to such an extent and to such a variety of performances as many curriculum planners at one time hoped and some still may hope. Moreover, the amount of transfer is not adequately distributed over a dichotomous variable; i.e., from one learning activity (A) to another or to any other performance (B) there might be marked transfer, but at the same time far less to another activity (C), although there might be transfer to both. This circumstance indicates why taxonomies of learning or the many psychometric classificatory systems of behavioral performances are at the same time numerous and dissatisfactory.

One might argue that in terms of learnable human performance--on which we narrow the further discussion--a psychometric taxonomy would be of broadest theoretical and practical value when it would allow for accurate prediction of transfer of learning. Instead, to date, transfer prediction research has rarely cared for the psychometric search of an economic and meaningful system to classify psychometric variables and vice-versa. Much of the transfer research has been conducted on the basic or micro level, trying different kinds of task analyses and different sorts of "similarity" measures as predictors (Skaggs, 1925; Robinson, 1927; Gibson, 1940; Boring, 1941; Osgood, 1949; Gagné, Baker, & Foster, 1950; Ahlstroem, 1961; Houston, 1964, 1965; Dallett, 1965; Butollo, 1968; for an extensive literature review see Flammer, 1970). This work has not allowed differential (psychometric) psychologists to make a connection to learning psychology. Some of the latest transfer-oriented taxonomic systems (Gagné, 1970, 1972; Klausmeier, 1971; Klausmeier, Ghatala, & Frayer, in press) could lend themselves better to psychometric "translations." At least the hierarchic conceptualization of learning activities has already proven to bear heavy implications on some traditional views of individual differences in school learning (Gagné & Paradise, 1962; Bloom, 1971; Flammer, 1973).

Psychometric taxonomies are traditionally based on correlations, typically in the factor analytic framework. Although they try to provide theoretical and generalizable insight (Spearman, 1923; Thurstone, 1938; Vernon, 1950; Meili, 1961; Guilford, 1967; and others), they are often used more as devices for economic descriptions of a given set of variables or of a set of individual test results. Because of a long-standing lack of interest in the learnability of the measured abilities or performances, the question for the adequacy of the systems to the prediction of transfer has rarely arisen, and when it has, there seems not to have been any doubt that transfer would very much depend on the correlation (Ferguson, 1954, 1956; Bunderson, 1964, 1967; Dunham, Guilford, & Hoepfner, 1966; Klauer, 1969; Fleishman, 1972).

Early Studies

A first empirical evaluation of the supposed connection was undertaken by Gengerelli (1934). He administered a code transcription task, first as a pretest, then as a posttest. In between, eight different groups of Ss were subjected to a three-minute interpolated activity, each group to another, varying in degree of similarity with the original activity.

The amount of retroactive inhibition was compared with the squared--uncorrected for

attenuation--coefficient of correlation. The retroactive inhibition was highest with high correlations and lowest with those in a middle range, increasing again as the correlation decreased.

A more recent attempt to investigate the connection between transfer and correlation was undertaken by Heinonen (1962). He hypothesized that the transfer of learning from test A to test B would be a linear function of the degrees of the angle between the two tests in the common factor space. This angle was expected to reflect the "functional similarity" relevant to transfer prediction and would be obtained through the equation

$$\cos \phi_{AB} = r_{AB}/(h_A h_B) ,$$

where h^2 = factor analytic commonality of each test. Heinonen assumed positive and negative transfer to be symmetric in the sense that with small angles there would be high positive or high negative transfer, depending on whether the kind of transfer was expected a priori to be positive or negative.

Heinonen tested his hypothesis, though only the positive transfer case, in an experiment with sensorimotor tests. Ninety-one 5th and 6th graders were presented with 16 tests on the first experimental day. On the four following days, for five minutes each day, they were retested on one of the 16 tests. Thereafter the 16 tests were readministered. The mean raw gains, divided by the pretest standard deviation of each test, were compared with the angles in the common factor space. After checking a graphic representation, Heinonen stopped the analysis, realizing "that the relationship between the gain due to training and the factorial distance between the training test and the other tests was not close."

With regard to the large, already existing research body on factorial classification of performance variables, such a finding would be alarming if it could be substantiated by further investigations. There are, however, at least two critical points that could account for the failure of a clear relation. The first might be the relatively short training time (four five-minute sessions); the other is the dependability of Heinonen's angle measure on the test sampling. The commonality is usually smaller than the reliability would allow it to be. If any test B happened to be a "factorial outsider," h_B^2 would tend to be low and $r_{AB}/(h_A h_B)$ would tend to be high. The mere inclusion of a very similar test B' into the test battery would raise h_B^2 and therefore lower $r_{AB}/(h_A h_B)$, although the amount of transfer per test would not be expected to change.

The correlation coefficient as such would not be affected by the composition of the eventual set of other tests. It would also be much more applicable, because it would not require the labor of administering a rather large number of additional tests.

Heinonen's hypothesis was subjected later to another experiment, done by Melametsä (1965). Melametsä used types of intelligence-test tasks and an experimental design that was the same as Heinonen's, except that he had, in addition, a control group working only on the pre- and posttests. The training he provided centered on an oddity task (number group test) and contained 12 sessions of ten tasks each.

Both the experimental and the control groups showed a significant mean change from pretest to posttest, but the experimental group exceeded the control group with statistical significance only in the learning task. Thus, further analysis of the relationship between transfer and $\cos \phi = r_{AB}/(h_A h_B)$ had to be abandoned.

Flammer (1970) hypothesized that the transfer would be a linear function of the squared coefficient of correlation after correction for attenuation, in order to protect the sign of the correlation coefficient:

$$T = a_0 + a_1 \frac{|r|^3}{r} .$$

The correlation for attenuation was intended to correct for the bias by which different coefficients are differently affected depending on the reliability of their variables. The division by the square root of the product of the reliability coefficients replaced the division by the square root of the product of the commonalities in the hypotheses by Heinonen and Melametsä. The argument for the squaring was to make the predictor express the relative amount of predictable individual difference variance.

An experiment was done including nine intelligence subtests selected from the BASC test battery (Cardinet & Rousson, 1967, 1968). These tests contained several series-continuation tasks, i.e., letter series, number series, geometric shape series, domino series, and date series, and some other tests like oddity problems, letter matrices, homophones, and numerical distance determinations.

The experimental design was the same as Heinonen's, except that the training consisted of ten 30-minute sessions, each on a different day, and there was a control group which did only the pretests and the posttests. The

experimental group consisted of 186 Ss and the control group of 235 Ss, all French-speaking 5th and 6th grade Swiss boys and girls. The training consisted of a systematically guided discovery of the principles underlying the kind of task in the letter-series test.

The transfer per test was measured as mean raw gain of the experimental group minus the mean raw gain of the control group, divided by the pretest standard deviation.

Table 1 contains the linear correlations r_{TC} between the predictor correlations and the transfer measures. The small number of transfer tests (eight) did not allow the smaller correlation coefficients to show up as statistically significant. The results may be summarized as follows:

1. There was a substantial linear component in the regression of the transfer per test on this correlation with the training test.
2. The correlations taken from the same group of Ss that also produced the transfer under investigation tended to be more predictive than those stemming from an unrelated group (cross-validation).
3. Posttest correlations tended to be more predictive than pretest correlations, as was the case in the cross-validation.
4. Squaring the predictor correlation coefficients had little and ambiguous effect.
5. The predictability increased after correction for attenuation, though very slightly. Since there also were only small differences in reliability, this effect could not show up more markedly.

TABLE 1
LINEAR CORRELATIONS r_{TC} BETWEEN TRANSFER MEASURES AND CORRELATIONS

	Predictors not corrected for attenuation		Predictors corrected for attenuation	
Predictors stemming from	not squared	squared	not squared	squared
Lxp. group/pretests	.43	.44	.54	.57
Exp. group/postests	.81*	.81*	.86*	.86*
Control group/pretests	.29	.22	.29	.19
Control group/posttests	.48	.46	.51	.49

*$p < .01$ (two-tailed); $df = 8 - 2 = 6$

II
A Conceptual Model

The connection between transfer and correlation has hitherto been conceived of rather intuitively. Flammer (1970) argued that the covariation between two performances within a population is an expression of the extent to which the two performances are psychologically or functionally the same or similar. Certainly, this correlational approach has the enormous advantage that it does not require detailed task analyses, let alone a S-R decomposition, and that it makes the large body of psychometric research available to learning theory and even to a wide range of instructional application. But the approach also has the disadvantage that correlation coefficients are highly sensitive to subject sampling.

The most important questions the author tries to clear up here are these:

1. How should the link between the mere statistical correlation measure and the transfer of learning be conceived of more coherently or more psychologically?
2. In a more technical respect, but relevant because of the weakness of the conceptualization to date, is it adequate to square the correlation coefficient or is it not? Regarding this, the empirical results so far are inconclusive. Are there stronger theoretical grounds to predict either the unsquared or the squared coefficient to be superior, i.e., stronger than the argument by analogy that the transfer would be proportional to individual difference variance components? Since this question taps at the same time the incertitude concerning the adequate scaling of the transfer, one would like to base the scaling question on a common theoretical ground, too.

The following model seems to have some ability to answer these questions. Assume a performance to be learned that consists of a number of elements whose extension corresponds to the average time required for a given population to master them. These elements are handled as hypothetical constructs and thought of as mutually exclusive and independent. Assume task X to be composed of n_x such elements and task Y of n_y elements. Let n_c be the number of elements which the given population of Ss identifies correctly as being common to both tasks and n_f be the number of elements identified incorrectly as being common to both tasks. Both n_c and n_f need to be regarded as being "identified" only insofar as they are learned.

Concerning the transfer, we assume now that learning X (so-called original learning) facilitates later learning Y or improves the performance Y to the extent that there are c-elements involved, which for their part have been affected by original learning, since they are included in X, too. If a proportion of k/n_x in X is learned (either k elements randomly selected and learned in the all-or-none fashion or all elements learned to the degree k/n_x), Y will profit to the extent n_c/n_y.

$$\Rightarrow \quad T_{k+} = \frac{k}{n_x} \frac{n_c}{n_y} \quad .$$

The analogue is true for elements falsely identified as being common. These elements produce a negative transfer component:

$$T_{k-} = \frac{k}{n_x} \frac{n_f}{n_y} \quad .$$

The net transfer is then

$$T_k = T_{k+} - T_{k-} = k \frac{n_c - n_f}{n_x n_y} \quad .$$

Concerning the product-moment correlation between X and Y, we propose to retain out of

the above proposed all-or-none and incremental learning assumptions only the all-or-none one. Incidentally, if in the incremental learning case the elements are conceived of as extremely small, the "error" introduced by dichotomizing the degree of learning to the alternative "learned" and "not yet learned" approaches zero. And since these elements are mere hypothetical constructs, nothing hinders this assumption.

The coefficient of the product-moment correlation between X and Y by definition is the following:

$$\rho = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\ \text{Var}(Y)}} \quad .$$

From the above assumption of the unrelatedness of the elements, Cov (X,Y) is in fact equal to the covariance produced by the c-elements and the f-elements, i.e.:

$$\text{Cov}(XY) = n_c \text{Cov}(c,c') + n_f \text{Cov}(f,f') \quad .$$

Since the correlation between two variables can never be different from zero if one or both variables have zero variance, a condition to evaluate a meaningful correlation coefficient is that both learning tasks X and Y are already learned to a certain degree, but not completely. We therefore define p as the degree of mastery of X and t as the degree of mastery of Y ($0 < p < 1$, $0 < t < 1$). Since for the c-elements and for the f-elements the probability to be mastered is the same as the probability to be mastered for all those elements of that task that has the higher degree of mastery, we define

$$v = \max(p,t) \quad .$$

Spelling out Cov (X,Y) produces:

$$\begin{aligned}\text{Cov}(X,Y) = {} & n_c(v(1-v)(1-v) + (1-v)(0-v)(0-v)) \\ & + n_f(v(1-v)(0-(1-v)) \\ & + (1-v)(0-v)(1-(1-v)))\end{aligned} \quad (1)$$

$$\begin{aligned}&= n_c(1-v)(v(1-v)+v^2) \\ &\quad + n_f(v(1-v)(v-1)+(1-v)(-v)v) \\ &= n_c v(1-v)(1-v+v) + n_f v(1-v)(v-1-v) \\ &= n_c v(1-v) + n_f v(1-v)(-1)\end{aligned}$$

$$\text{Cov}(X,Y) = (n_c - n_f)v(1-v)$$

Since $\text{Var}(X) = n_x p(1-p)$ and $\text{Var}(Y) = n_y t(1-t)$

$$\rho = \frac{(n_c - n_f)v(1-v)}{\sqrt{n_x p(1-p) n_y t(1-t)}} \quad .$$

Define $w = \min(p,t)$,

$$\Rightarrow \rho = \frac{n_c - n_f}{\sqrt{n_x n_y}} \sqrt{\frac{v(1-v)}{w(1-w)}} \quad .$$

For the special case that $p = t$,

$$\rho = \frac{n_c - n_f}{\sqrt{n_x n_y}} \quad .$$

We remember

$$T_k = \frac{n_c - n_f}{n_x n_y} k \quad .$$

Substituting by ρ we get

$$T = \frac{k}{\sqrt{n_x n_y}} \rho \quad .$$

This also makes intuitive sense: The transfer is directly related to the amount k of so-called original learning, but it becomes small if either the original or the transfer task or both become "bigger," since we defined T from the beginning as the amount of facilitation or interference relative to the amount of learning required.

If $p \neq t$

$$T_k = \rho \frac{k}{\sqrt{n_x n_y}} \sqrt{\frac{w(1-w)}{v(1-v)}} \quad .$$

Adding the unity factor

$$\sqrt{\frac{n_v n_w}{n_w n_v}} \quad ,$$

(1) Here it is assumed that all the f-elements are "right" in the first task and "wrong" in the second task. The result would not be changed if the contrary were true or if there would be a mixture (some "right" in the first task and some in the second).

where $n_v = n_x$, if max (p,t) = p

$= n_y$, if max (p,t) = t

$n_w = n_x$, if min (p,t) = p

$= n_y$, if min (p,t) = t,

gives

$$T_k = \rho \frac{k}{\sqrt{n_x n_y}} \sqrt{\frac{n_v}{n_w}} \sqrt{\frac{n_w w(1-w)}{n_x v(1-v)}}$$

$$= \rho \frac{k}{\sqrt{n_x n_y}} \sqrt{\frac{n_v}{n_w}} \frac{\sigma_w}{\sigma_v} .$$

Since either ($n_x = n_v$ and $n_y = n_w$) or ($n_x = n_w$ and $n_y = n_v$),

$$T_k = \rho \frac{k}{n_w} \frac{\sigma_w}{\sigma_v} . \quad (2)$$

One important characteristic of this conceptualization is that for a given learning time k, the transfer from the original learning to Y is the same as from Y to X. This is given by the definition of the amount of transfer relative to the amount of the total time needed to learn a given task completely and from zero. An intuitive reason for this is best demonstrated by some examples, arranged in Table 2. One could paraphrase as follows: when $n_x < n_y$, X, with a given $(n_c - n_f)$, produces a large "carryover" which in turn loses relative weight in the large Y. If, on the other hand, $n_x > n_y$ with the same $(n_c - n_f)$, the carryover is smaller but its weight within Y is heavier.

Table 2 shows that the transfer formula is also applicable to the special cases of "simple learning" when either $n_x = n_c$, or $n_y = n_c$, or $n_x = n_c = n_y$.

Coming back to the question of whether the transfer is proportional to the squared or to the simple correlation coefficient, it is evident that within the proposed system of conceptualizing both the transfer and the correlation at once, only the unsquared correlation coefficient makes sense. Moreover, the relation should hold even if $\rho < 0$.(3)

Concerning the width of application of the model a crucial question seems to be that of the definition of the task. The model itself requires the single c-elements and the f-elements to be learned at the same rate as all other elements. This can theoretically be accomplished by having all single elements having the same probability to be sampled or in a case of a fixed sequence of elements by having the c-elements and the f-elements each arranged at equal intervals, i.e., n_c/n_x and n_f/n_x, respectively. Further research must determine how critical these conditions are and how their being met or not met can be decided objectively.

(2) The interpretation of this result is complicated, since growth of v increases σ_v as long as $v < .5$, but decreases σ_v thereafter. Furthermore, it can happen that $v > .5$ and $w < .5$, by which alone it could even not be decided whether $\sigma_w/\sigma_v > 1$ or $\sigma_w/\sigma_v < 1$ or $\sigma_w/\sigma_v = 1$.

(3) One might remember that Heinonen (1962) in fact expected negative transfer, provided it was negative for whatever reasons, to be higher (negative) the bigger the (cor-)relation as expressed in terms of narrowness in the common factor space.

TABLE 2
ILLUSTRATIVE EXAMPLES OF TRANSFER OF LEARNING

n_x	n_y	$n_c - n_f$	k	$\frac{k}{n_x}$	$\frac{n_c - n_f}{n_y}$	$T_k = k \frac{n_c - n_f}{n_x n_y}$
5	10	3	2	2/5	3/10	3/25
10	5	3	2	2/10	3/5	3/25
5	5	3	2	2/5	3/5	6/25
5	3	3	2	2/5	3/3	2/5
3	5	3	2	2/3	3/5	2/5
5	5	5	2	2/5	5/5	2/5
5	10	-3	2	2/5	-3/10	-3/25

III
Validation Experiment[4]

Since the experiments done by Gengerelli (1934), Heinonen (1962), and Melametsä (1965) were not conclusive, the affirmative results of the Flammer (1970) experiment needed a replication. In preparing this the following improvements of the design were effectuated at the same time:

1. Since the posttest correlations seemed to be more predictive, a multitude of experimental groups with different learning times should manifest a region of degree of learning which would yield the optimally predictive correlation coefficient.
2. In the former experiment the reliabilities of the variables were below psychometric standards for reasons of drastic testing time reductions. This could have been the cause for the lack of statistical significance of the prediction in using either experimental or control group pretest correlations. Adding and substituting some transfer tests and reextending the testing times should yield higher reliabilities.
3. Since the squaring of the predictor correlations gave equivocal results, a broader range of the predictor correlations should allow for clearer effects. Broadening the range to an interval from $r = 0.0$ to $r = .65$, say, was also expected to raise the (relative) predictive value.

Why should it be that posttest correlations are more predictive than pretest correlations? Although the correction for attenuation does eliminate the bias due to unreliability and, for a comparison of different tasks, compensate for unequal lacks of reliability, this correction never can raise the reliability itself. The opposite is true: the standard error becomes even larger. It is therefore desirable in all cases to use variables which are as reliable as possible. It is common experience (Tinkelman, 1971, p. 63) that the reliability of tests is best when the item difficulty indices are around $\underline{p} = .5$ or even a bit higher (Mikkonen, 1972), since the probability of guessing is larger with difficult items than with easy ones. In the former experiment as in the one to be reported now, for reasons of trainability tests were chosen which were originally prepared for and normally administered to older students. Therefore it could be that the average $\underline{p}$'s came closer to .5 through training. And the replication experiment that included various lengths of learning times could therefore eventually show a decrease of predictive validity after an increase, namely after passing the point $\underline{p} = .5$.

Subjects. Three hundred and eight-five German-speaking Swiss fifth graders were equally distributed among three experimental and one control group--E3, E6, E9, and C.

Learning task. All the training centered on tasks of the type of the test LS (letter series to continue), the same as in the earlier experiment.

Transfer tasks. Twelve paper-and-pencil tests were chosen from Cardinet and Rousson (1967, 1968), Amthauer (1963), and Flammer et al. (1971). Again, there were several series continuation tests with different materials such as dominoes, geometric shapes, numerals in series, numerals in matrices, and ornaments. Other tests were less similar to the learning test, like analogies, short-term memory, and mathematical operations tests. All the tests

[4] Parts of the following results were prepared for the diploma-theses of Flühler (1971), Kramis (1971), and Stöckli (1972), together with some analyses not covered here.

were rather difficult for the selected Ss. To depress further the probability of ceiling effects, all the normal testing times were reduced by one-fifth.

To the extent that the Ss had different mastery levels on either the learning test and the transfer tests or on different transfer tests, the dependency of the transfer on the intercorrelation was expected to be altered according to the model proposed here. The same thing would be true for different "sizes" of the tasks. It is probable that these differences are rather slight, at least in the pretests. The following analysis will not take into account these eventual differences and therefore will represent a (statistically conservative) check of just the general features of the model. The main reason for this is that the practical evaluation of the n's and the p's would require a learning time scale. Such a scale would first have to be effectuated experimentally--a very time-consuming procedure.

Experimental training. The instruction was provided exclusively through booklets of Skinner-type linear learning programs prepared specially for this experiment. The Ss had to learn how the letter series were constructed by using specially devised graphical means. All experimental Ss were presented with each kind of problem in the letter-series test. Those with extended training received more and increasingly difficult examples.

Transfer scaling. All the pretest distributions were normalized, and the scale yielded was used to determine the transfer for each test, which was defined as mean gain of the respective experimental group minus the mean gain of the control group.

Again, the model proposed here would have required learning time scales. Since the substitute chosen in this experiment can easily be effectuated, it would be very welcome if it would work satisfactorily.

Hypotheses

The general hypothesis was (1) that the regression of the transfer measures on the unsquared coefficient of intercorrelation after correction for attenuation would contain a significant linear component and nothing but a linear component. (2) The predictive value should increase with the lengthening of learning time and eventually decrease after a certain point, for the reasons of reliability mentioned above. Certainly, according to the finally given formula for the transfer depending on ρ, n_w, σ_w, and σ_v, at the same time a regression deteriorating effect was expected with the growing of the differences between the variance of the transfer tests (presumably σ_v). (3) The regression curve according to the model should yield no significant intercept, i.e., $E(T \mid r = 0) = 0$.

Results

The prediction equations contained all 12 transfer tests with their average transfer measures and their intercorrelations with the training test LS. Corresponding to the model, LS itself could have been included as a special transfer case, where $n_x = n_y = (n_c - n_f)$. The "intercorrelation" would have been equal to the reliability, and after correction for attenuation would equal the value 1.00. Since it was felt that relatively high learning effects on the training test were to be expected in any case, independently from the specific features of the model, it would not have been fair to include the learning test in the evaluation of the regression equations. Incidentally, some exploratory analyses including the learning tests have been made: the results showed indeed much stronger overall relationships than the ones presented here.

Since for practical reasons it was impossible to raise the number of transfer tests to more than 12, the degrees of freedom for the prediction equation were rather low. A first trend analysis was therefore made graphically and by optical inspection. None of the scatter plots of the many possible combinations showed any tendency toward a nonlinear trend when the predictor correlations were unsquared (Flüh1er, 1971). The further exposition of the results can therefore be restricted to the linear correlations r_{TC} as indices for predictive validity. For reasons which follow the question concerning the squaring of the correlation coefficients is postponed.

The reliability coefficients for the unusual S-population and the reduced testing times were estimated as retest correlations of the control group. They varied between .40 and .75. Correction for attenuation, not surprisingly, improved every prediction. It seems reasonable to include here only the results for which the correction for attenuation was done.

Table 3 contains the r_{TC}. With very few exceptions the r_{TC}'s are statistically significant, even in the cross-validation cases. Some coefficients are of an impressive height, i.e., in the .80's.

Apparently the transfer of E6 was best predicted with only one exception. At the same time the posttest intercorrelations were the best predictors, among them the ones stemming

TABLE 3
CORRELATIONS BETWEEN PREDICTOR-CORRELATIONS AND TRANSFER MEASURES

Predictor stemming from		Predicted transfer in E3	E6	E9	Average over all three groups
Pretests	C	.58*	.68*	.57*	.61
	E3	.54*	.63*	.61*	.59
	E6	.62*	.80**	.75**	.73
	E9	.57*	.59*	.47	.55
Average over pretests ($\bar{r}_{TC}$)		.58	.69	.61	.63
Posttests	C	.57*	.63*	.48	.57
	E3	.81**	.90**	.82**	.85
	E6	.70**	.88**	.86**	.83
	E9	.63*	.58*	.47	.57
Average over posttests ($\bar{r}_{TC}$)		.69	.79	.70	.73
Average over pre- and posttests ($\bar{r}_{TC}$)		.64	.74	.66	.68

$\bar{r}_{TC}$ averaged on Fisher's z-scale

*$p < .05$ (one-tailed)

**$p < .01$ (one-tailed)

from E3 and E6. These differences, however, did not reach statistical significance, since the mean square between all the 24 corresponding Fisher's z's divided by the predefined error variance of z ($= 1/df = 1/9$), yielded $F < 1$. All further statistical analysis of the 2 (pre/posttests) x 4 (intercorrelations out of C, E3, E6, and E9) x 3 (transfer in E3, E6, and E9) design seemed unadvisable after this.

A puzzling characteristic of the results is that E9 not only yielded the worst predictors and transfer that was less predictable than E6, but the worst predictions of all were those to which E9 furnished both transfer and predictors. Several ceiling checks disconfirmed the ceiling hypothesis, be it for the learning-test raw scores or the 12 transfer-test raw scores, or even the transfer variance between the transfer tests. Since all the testing and training was done under the members of the research team, any manifest irregularity in this could be excluded, except for the observation that the teachers of the E9 classes towards the end of the experiment began to be discontent with the amount of instruction time being lost because of the investigation. The teachers had been informed in advance of the exact times, but some had apparently underestimated the plan. All that can be said in regard to E9 is that eventually the teachers' attitudes may have confused the Ss who in turn may then have taken the experiment less seriously.

The model's prediction that the transfer would linearly depend on the unsquared correlation coefficient and not on the squared one was confirmed insofar as the r's yielded better predictions than the r^2's in 13 out of the 14 cases where E9 was involved neither in the predictors nor in the transfer measures. The 13 cases involving E9 showed the r's to be superior only five times. By disregarding the rather puzzling E9 group, the model's prediction is to be considered as clearly confirmed.

The third hypothesis was concerned with the intercept of the linear regression equations expected to be zero. Table 4 shows that they were negative without exception. The largest departures were those of the cases which already had shown the most precise predictions. The third hypothesis is clearly to be rejected.

TABLE 4
t-VALUES FOR DEPARTURE FROM ZERO OF THE
INTERCEPTS OF THE LINEAR REGRESSION CURVES

Predictor stemming from		Predicted transfer in E3	E6	E9
Pretests	C	- .72	-1.93	-1.05
	E3	- .45	-1.56	-1.18
	E6	-1.10	-3.17**	-2.36*
	E9	- .73	-1.46	- .63
Posttests	C	- .58	-1.56	- .56
	E3	-2.56*	-5.30**	-3.08*
	E6	-1.61	-4.60**	-3.89**
	E9	- .97	-1.37	- .62

*$p < .05$ (two-tailed)

**$p < .01$ (two-tailed)

GPO 803-822-3

IV
Discussion

In general, the confirmed predictions are not only statistically significant, but also of appreciable accuracy. This is true when predictors and transfer measures are from the same sample, and in the cross-validation case as well. Clearly the transfer is proportional to the unsquared correlation coefficient, not to the squared one.

In opposition to what was to be expected from the model, the regression curve intercepts were negative. This had already been found in the 1970 experiment.

Also, in both experiments there were a few tests with negative transfer measures. According to the model this can only happen if $n_f > n_c$. But under the model's assumptions this would at the same time cause the correlation to be negative, and no non-zero intercept is justifiable by this. Negative intercepts need then to be caused by either an amount of negative transfer on which Ss do not vary or by some additional component of positive covariation which is not related to transfer.

The first explanation does not seem to be too risky to this author. Standardized tests and especially experiments like the one discussed here are rather exceptional events in Swiss schools. Thus, the unusual circumstances and the unusual emphasis of the experimental training may well have led all the experimental Ss to invariably overgeneralize the learned kind of solution approach to all the other posttests. If this were proven to be true, the negative intercept would be typical only for either the specially emphatic training or the S sample or both.

The second possible explanation that the author sees could call for a slight modification of the model. It could be argued that the covariation between two tasks is the sum of two components, the one described by the model thus far and one produced by individual differences in overall information processing rate, like personal speed, attitude toward intellectual performance, and engagement in the specific kind of experimental setup. At least for this length of training it seems plausible that such a covariation would not have noticeable transfer implications. To test this explanatory hypothesis, one could statistically hold constant some general performance variable like mental age.

While both of these proposed explanations could account for negative regression intercepts, the first is much more likely because of the fact that some tests showed a negative transfer, which may be regarded as rather exceptional for the kind of cognitive tasks used.

One thing this and the former experiment showed in comparison with the ones by Heinonen (1962) and Melametsä (1965) is that transfer studies with psychometric tasks, which usually are chosen for their presumed stability among other characteristics, need to be done with rather extensive and potent instruction. This may be prohibitive for researchers and for the Ss or the persons who decide whether Ss are available.

Certainly the two former experiments also worked with the factor analytic angles as predictors and not with the correlations. Reanalysis of Heinonen's data yielded r_{TC}'s as follows: -.04 (pretest intercorrelations) and .39 (posttest intercorrelations), both corrected for attenuation. The second r_{TC} is at least encouraging. It might have been higher had there been a control group allowing subtraction of learning and transfer effects not attributable to the experimental training. The data by Melametsä are not reanalyzable in this context, since the intercorrelations were not reported.

The proposed model is confirmed with the presented data although several of its implications still have not been investigated empirically. Particularly, these are the variations in $p - t$ and $n_x - n_y$, and also the negative transfer case. It is highly desirable to have a set of tasks with an empirically normed learning time scale. This scale would allow identification of p and t as well as n_x and n_y--for the latter

two at least their relative values, which are the only ones that are needed. n_c and n_f never enter into a formula relating transfer to correlation.

Finally it might be realized that the proposed approach as such is not yet concerned with individual differences in transfer, but uses the existence of individual differences in performance in order to get predictors for the average transfer. By this the differential (correlational) approach and the general (experimental) approach (Cronbach, 1957) are neither mutually exclusive nor even independent and separate from one another.

References

Ahlstroem, K. G. Transfer and Communality: An Experimental Study of Learning. Uppsala: Almqvist and Wiksells, 1961.

Amthauer, R. Intelligenzstruktur-Test. Goettingen: Hogrefe, 1963.

Bloom, B. S. Individual differences in school achievement: A vanishing point? Paper presented at the AERA Convention, 1971.

Boring, E. G. Communality in relation to proaction and retroaction. American Journal of Psychology, 1941, 54, 280-283.

Bunderson, C. V. Transfer functions and learning curves: The use of ability constructs in the study of human learning. Research Bulletin 64-62. Princeton, N.J.: Educational Testing Service, 1964.

Bunderson, C. V. Transfer of mental abilities at different stages of practice in the solution of concept problems. Research Bulletin 67-20. Princeton, N. J.: Educational Testing Service, 1967.

Butollo, W. Subjektive und objektive Aehnlichkeit im verbalen Lernen. Dissertation, Wien, 1968.

Cardinet, J., & Rousson, M. Etude factorielle de tests d'aptitude scolaires. Revue Suisse de Psychologie, 1967, 26, 256-270; 1967, 26, 362-380; 1968, 27, 40-66.

Cronbach, L. J. The two disciplines of scientific psychology. American Psychologist, 1957, 12, 671-684.

Dallett, K. M. A transfer surface for paradigms in which second-list S-R pairings do not correspond to first-list pairings. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 528-534.

Dunham, J. L., Guilford, J. P., & Hoepfner, R. Abilities pertaining to classes and the learning of concepts. Report No. 39. Los Angeles: Univ. of Southern California, 1966.

Ferguson, G. A. On learning and human ability. Canadian Journal of Psychology, 1954, 8, 95-112.

Ferguson, G. A. On transfer and the abilities of man. Canadian Journal of Psychology, 1956, 10, 121-131.

Flammer, A. Transfer und Korrelation. Weinheim: Beltz, 1970.

Flammer, A. Does computer-assisted instruction reduce individual differences? Stanford University: Institute for Mathematical Studies in the Social Sciences, Technical Report, 1973 (in press).

Flammer, A., Broch, E., Bründler, J., Imfeld, J., Kramis, J., & Niedermann, A. AzN--Aufgaben zum Nachdenken, Schweizer Bearbeitung. Basel: Beltz, 1971.

Fleishman, E. A. On the relation between abilities, learning, and human performance. American Psychologist, 1972, 27, 1017-1032.

Flühler, R. Korrelation als Transferprädiktor. Unpublished Diploma-Thesis. University of Fribourg/Switzerland: Psychologisches Institut, 1971.

Gagné, R. M. The Conditions of Learning. (2nd ed.) New York: Holt, Rinehart and Winston, 1970.

Gagné, R. M. Domains of learning. Interchange, 1972, 3, 1-8.

Gagné, R. M., Baker, K. E., & Foster, H. On the relation between similarity and transfer of training in the learning of discrimination motor tasks. Psychological Review, 1950, 57, 67-79.

Gagné, R. M., & Paradise, N. E. Abilities and learning sets in knowledge acquisition. Psychological Monographs, 1962, 75, 14 (Whole No. 518).

Gengerelli, J. A. Similarity and retroaction. Journal of Experimental Psychology, 1934, 17, 680-690.

Gibson, E. J. A systematic application of generalization and differentiation to verbal learning. Psychological Review, 1940, 47, 196-229.

Guilford, J. P. The Nature of Human Intelligence. New York: McGraw-Hill, 1967.

Heinonen, V. A factor analytical study of

transfer of training. Scandinavian Journal of Psychology, 1962, 3, 177-188.

Houston, J. P. Verbal transfer and interlist similarities. Psychological Review, 1964, 71, 412-414.

Houston, J. P. A repaired S_1-R_2 and S_2-R_1 transfer surface. Psychonomic Science, 1965, 3, 343-344.

Klauer, K. J. Lernen und Intelligenz. Weinheim: Beltz, 1969.

Klausmeier, H. J. Cognitive operations in concept learning. Division 15 Presidential Address, APA Convention, 1971. Educational Psychologist Newsletter, November 1971.

Klausmeier, H. J., Ghatala, E. S., and Frayer, D. A. Conceptual Learning and Development: A Cognitive View. New York: Academic Press, in press.

Kramis, J. Transfer als Strukturänderung. Eine faktorenanalytische Untersuchung über die Veränderung der Intelligenzstruktur infolge Trainings. Unpublished Diploma-Thesis. University of Fribourg/Switzerland: Psychologisches Institut, 1971.

Meili, R. Lehrbuch der psychologischen Diagnostik. Bern: Huber, 1961.

Melametsä, L. The influence of training on the level of test performance and the factor structure on intelligence tests. Scandinavian Journal of Psychology, 1965, 6, 19-25.

Mikkonen, V. Changes in test characteristics as a function of learning of the object-task. Scandinavian Journal of Psychology, 1972, 13, 306-313.

Osgood, C. E. The similarity paradox in human learning: A resolution. Psychological Review, 1949, 56, 132-143.

Robinson, E. S. The similarity factor in retroaction. American Journal of Psychology, 1927, 39, 297-312.

Skaggs, E. B. Further studies in retroactive inhibition. Psychological Monographs, 1925, 34, No. 161.

Spearman, C. The Nature of Intelligence and the Principles of Cognition. London: Macmillan, 1923.

Stöckli, H. Transferverlauf. Unpublished Diploma-Thesis. University of Fribourg/Switzerland: Psychologisches Institut, 1972.

Thurstone, L. L. Primary mental abilities. Psychometric Monographs, 1938.

Tinkelman, S. N. Planning the objective test. In R. L. Thorndike (Ed.), Educational Measurement. Washington, D. C.: American Council on Education, 1971. Pp. 46-80.

Vernon, P. E. The Structure of Human Abilities. New York: Wiley, 1950.

GPO 803-822-2

National Evaluation Committee

Helen Bain
Immediate Past President
National Education Association

Lyle E. Bourne, Jr.
Institute for the Study of Intellectual Behavior
University of Colorado

Jeanne S. Chall
Graduate School of Education
Harvard University

Francis S. Chase
Department of Education
University of Chicago

George E. Dickson
College of Education
University of Toledo

Hugh J. Scott
Superintendent of Public Schools
District of Columbia

H. Craig Sipe
Department of Instruction
State University of New York

G. Wesley Sowards
Dean of Education
Florida International University

Benton J. Underwood
Department of Psychology
Northwestern University

Robert J. Wisner
Mathematics Department
New Mexico State University

Executive Committee

William R. Bush
Director, Program Planning and Management
Deputy Director, R & D Center

Herbert J. Klausmeier, Committee Chairman
Principal Investigator
R & D Center

Joel R. Levin
Principal Investigator
R & D Center

Donald J. McCarty
Dean, School of Education
University of Wisconsin

Richard A. Rossmiller
Director
R & D Center

Dan Woolpert
Director, Management Systems
R & D Center

Faculty of Principal Investigators

Vernon L. Allen
Professor
Psychology

Frank H. Farley
Associate Professor
Educational Psychology

Marvin J. Fruth
Associate Professor
Educational Administration

John G. Harvey
Associate Professor
Mathematics

Frank H. Hooper
Associate Professor
Child Development

Herbert J. Klausmeier
V. A. C. Henmon Professor
Educational Psychology

Stephen J. Knezevich
Professor
Educational Administration

Joel R. Levin
Associate Professor
Educational Psychology

L. Joseph Lins
Professor
Institutional Studies

James Lipham
Professor
Educational Administration

Wayne Otto
Professor
Curriculum and Instruction

Robert Petzold
Professor
Curriculum and Instruction

Thomas A. Romberg
Associate Professor
Curriculum and Instruction

Richard A. Rossmiller
Center Director
Professor, Educational Administration

Richard L. Venezky
Associate Professor
Computer Science

Alan M. Voelker
Assistant Professor
Curriculum and Instruction

Larry M. Wilder
Assistant Professor
Communication Arts